Proceedings of

2nd ICSIIT 2010

International Conference on Soft Computing, Intelligent System and Information Technology

1-2 July 2010, Bali, Indonesia

ICSIIT

Supported by

MERATUS

APTIKOM

ASOSIASI PERGURUAN TINGGI INFORMATIKA DAN KOMPUTER

IBM

Proceedings

ICSIIT 2010

International Conference on
Soft Computing, Intelligent System and Information Technology

1-2 July 2010

Bali, Indonesia

Editors:
Leo Willyanto Santoso
Andreas Handojo

Informatics Engineering Department
Petra Christian University

Center of Soft Computing and
Intelligent System Studies

Proceedings

International Conference on Soft Computing, Intelligent System and Information Technology 2010

Publisher:
Informatics Engineering Department
Petra Christian University

ISBN: 978-602-97124-0-7

Additional copies may be ordered from:
Informatics Engineering Department
Petra Christian University, Siwalankerto 121-131, Surabaya 60236, Indonesia

Cover Art production by Adi Wibowo/Informatics Engineering Department

ICSIIT 2010
Table of Contents

Artificial Intelligence & Applications

Knowledge & Data Engineering

Imaging Technology

Internet, Web Services & Mobile Applications

Communication Systems & Networks

Control & Automation

Preface

First of all, I would like to give thank to God the Creator, God the Redeemer and God who leads us to the truth for all His blessings to us. As we all know, this 2nd International Conference on Soft Computing, Intelligent Systems and Information Technology 2010 (ICSIIT 2010) is held from 1-2 July 2010 in the Hard Rock Hotel located at this paradise island, Bali, Indonesia. I thank Him for His presence and guidance in letting this conference happen. Only by God's grace, we hope we could give our best for 2nd ICSIIT 2010 despite of all of our limitation.

We have received more than 130 papers from 15 countries. Only 96 papers from 13 countries have been accepted based on reviewers' ratings and comments. The paper selection process was based on full paper submissions. We thank all authors who have contributed and participated in presenting their works at this conference. We also gratefully acknowledge the important review supports provided by the 19 members of the program committee from 8 different countries. Their efforts were crucial to the success of the conference.

We are also so blessed by the presence of two invited speakers who will address the important trends relating to natural languages processing and soft computing. The first issue on natural language will be addressed by a lovely professor, Prof. Rachel Edita O. Roxas, Phd. who will present "Human Language Technology: the Philippine Context". We are aware that the main problem in language processing is ambiguity from syntax level to semantic level. In my personal opinion, we are also living in between inherently ambiguous and completely reasonable world. Einstein once said that "As far as the laws of mathematics refer to reality, they are not certain, as far as they are certain, they do not refer to reality." Prof. Rolly Intan, Dr.Eng will address this issue on soft computing with his presentation entitled "Mining Multidimensional Fuzzy Association Rules from a Normalized Relational Database".

I hope during your stay in this beautiful island you will enjoy and benefit both, the fresh sea breeze and harmonious sound from sea waves, as well as the intellectual and scientific discussions. I hope your contributions and participation of the discussion will lead to the benefit of the advancements on Soft Computing, Intelligent Systems and Information Technology.

Soli Deo Gloria,
Iwan Njoto Sandjaja
Conference Chair
ICSIIT 2010 Bali Indonesia

Organizing Committee

The first ICSIIT 2010 is organized by Informatics Engineering Department, in cooperation with the Center of Soft Computing and Intelligent System Studies, Petra Christian University, Indonesia.

Conference Chair:

Iwan Njoto Sandjaja	Petra Christian University, Indonesia
Adi Wibowo	Petra Christian University, Indonesia

Organizing Committee:

Agustinus Noertjahyana	Petra Christian University, Indonesia
Alexander Setiawan	Petra Christian University, Indonesia
Andreas Handojo	Petra Christian University, Indonesia
Djoni Haryadi Setiabudi	Petra Christian University, Indonesia
Gregorius Satia Budhi	Petra Christian University, Indonesia
Ibnu Gunawan	Petra Christian University, Indonesia
Justinus Andjarwirawan	Petra Christian University, Indonesia
Kartika Gunadi	Petra Christian University, Indonesia
Leo Willyanto Santoso	Petra Christian University, Indonesia
Liliana	Petra Christian University, Indonesia
Lily Puspa Dewi	Petra Christian University, Indonesia
Rudy Adipranata	Petra Christian University, Indonesia
Silvia Rostianingsih	Petra Christian University, Indonesia
Yulia	Petra Christian University, Indonesia

Program Committee

Grant Pogosyan (Japan)

Jan Chan (Australia)

Kevin Wong (Australia)

Kwan Pyo, Ko (Korea)

Masao Mukaidono (Japan)

Moeljono Widjaja (Indonesia)

M. Rahmat Widyanto (Indonesia)

Nelson Marcos (Philippines)

Masashi Emoto (Japan)

Noboru Takagi (Japan)

Rolly Intan (Indonesia)

Budi Bambang (Indonesia)

Rudy Setiono (Singapore)

Shan Ling, Pan (Singapore)

Son Kuswadi (Indonesia)

Tae Soo, Yun (Korea)

Xiuying Wang (Australia)

Yung Chen, Hung (Taiwan)

Zuwairie Ibrahim (Malaysia)

Search Engine Application Using Fuzzy Relation Method for E-Journal of Informatics Department Petra Christian University

Leo Willyanto Santoso
Petra Christian University
Siwalankerto 121-131
Surabaya 60236 INDONESIA
+6231-2983455
leow@petra.ac.id

Rolly Intan
Petra Christian University
Siwalankerto 121-131
Surabaya 60236 INDONESIA
+6231-2983455
rintan@petra.ac.id

Prayogo Probo Susanto
Petra Christian University
Siwalankerto 121-131
Surabaya 60236 INDONESIA
+6231-2983455

ABSTRACT
Nowadays, scientific articles are easily obtained, because many researchers who conduct research discover new things. However, the increasing number scientific articles is not accompanied by the availability of applications to assist in the search for relevant articles. Today, available search engine applications perform only a search process based on string matching of search terms. In this research, a search engine application based on keyword relevance by using fuzzy relationship was developed.

This search engine application is built using PHP programming language and mysql as its database. Windows XP is used as the operating system. The used methods in fuzzy relationship are keyword used to paper, paper to paper, and paper to keywords and keyword to keyword. In addition, the components used to convert pdf files into plain text format.

Based on the results of experiments conducted, the process of searching for the 25 articles takes less than 5 seconds. For the indexing process, it is influenced by the number of pages per article.

Keywords
Fuzzy Relation, Search Engine, Paper.

1. INTRODUCTION
At this time, journal is one of the many forms of documents selected by the researchers and scientists to put the results of experiments or research that have been conducted. Through the journal, the researchers poured all aspects of the research that been conducted by attaching a detailed information about the research he had conducted.

A journal as a medium of information from the experts/researchers to the public media has an important role and very strategic. For example in the field of education, the journal serves as a good material for teaching materials (for teachers) or as a reference for students to learn a new science.

Currently, website has been highly developed. This resulted in a website used as a medium of publication of a journal from experts/researchers to the public or known by the name of the e-journal. But more and more of the e-journal are not followed by the use of search engines technology. The search engine on each of the e-journal that is useful to facilitate a user who wants to do a search on a journal and other journal/articles that may still relate to one another is needed.

The problem is how to design search engines on e-journal that can produce a related mutually journal to one another based on the keywords that are input by the user.

This paper presents a new search engine applications that do not only search on the similarity keyword provided by journal or scientific paper, but also provide a reference paper which relate to each other as well as journal is desired by the user.

The remaining part of this paper is organized as follows. Section 2 presents an overview of current proposal for dealing with fuzzy relation. Section 3 depicts the approach that we have delineated to solve the proposed problems. Section 4 discusses the performance of proposed methods. Finally, section 5 concludes the paper.

2. FUZZY RELATION
Fuzzy relation is a method for explaining the relationship of two different things (completely different). As illustration, the word "apple" (apple) and "tiger" (tiger) then in general the two words are not related. In general, the word "apple" refers to the name of the fruit and the "tiger" refers to the name of wild animal.

In the computer world there is manufacturer software, Macintosh (Mac). Mac has the brand "apple" so often referred to as the Apple Macintosh. Recently, Mac issued a new operating system called "Tiger" OS. From the relationship with Mac as the word "Apple" and "Tiger" is actually not related in general and in writing, have a relationship in the world of computers. Given the fuzzy relation then this kind of relationship will be examined with an assumption and goal that by knowing the relationship closeness/kinship between the two word/object. In relation to the world of search (searching), then by inserting the word "apple", there is the possibility of the word tiger will also be a result of output. Not because the results wrong, but because between the word "apple" and "tiger" there is kinship [4].

Explanation of fuzzy relation can also be described as follows: two words that completely unrelated (eg: "apple" and "tiger"), will have a relationship when both the word is addressed in one document. More and more documents that discuss both the relationship between the two words ("apple" and "tiger") will be getting closer.

Fuzzy Relation will search 4 links from a combination of words (keywords) and documents (paper) these relationships are:

- *Keyword* to *paper*
- *Paper* to *paper*
- *Paper* to *keyword*
- *Keyword* to *keyword*

Explanation of each of this relationship along with the calculation process is described as follows:

1. At this step is assumed relationship between the keyword to the weight of the paper has value to the paper of the following keywords:

   $P = \{P_1, P_2, \ldots Pn\}$ is a *set of papers*

   $D = \{D_1, D_2, \ldots Dn\}$ is a *set of keywords*

   For example from the data obtained by paper and keyword relationship expressed as a fuzzy set of papers on the following keywords:

   $P_1 = \{0.3/D_2, 0.7/D_5, 1/D_7, 1/D_8\}$,

   $P_2 = \{1/D_2, 0.8/D_5, 0.8/D_7, 1/D_8\}$,

   $P_3 = \{0.9/D_1, 0.9/D_3, 1/D_4, 0.8/D_6\}$,

   $P_4 = \{1/D_1, 0.5/D_3, 0.8/D_4, 0.8/D_6\}$,

   $P_5 = \{0.1/D_2, 0.7/D_5, 1/D_4, 1/D_8\}$,

   $P_6 = \{0.9/D_2, 1/D_5, 0.8/D_4, 1/D_8\}$

   For $P = \{P_1, P_2, \ldots P_6\}$ and $D = \{D_1, D_2, \ldots D_8\}$, where each *paper*/document regarded as a fuzzy set of keywords so we get $\mu P_5 (D_1) = 0.1$.

2. *Similarity* between 2 *papers* expressed as a function of R where $R: P \times P \rightarrow [0.1]$ [5]

$$R(P_i,P_j) = \frac{\sum_D (\mu_{Pi}(D), \mu_{Pj}(D))}{\sum_D \mu_{Pj}(D)} \qquad (2.1)$$

   Where:

   R: Relation

   P_i: *Paper*/document i

   P_j:*Paper*/document j

   D: *Keyword*

   μ: *Membership function* as a *mapping* $\mu P_i: D \rightarrow [0,1]$.

   Can find a relationship between a paper with one another, eg:

$$\text{relation } R(P_1,P_2) = \frac{0.3+0.7+0.8+1}{1+0.8+0.8+1} = \frac{2.9}{3.6} = 0.78$$

   The calculation of paper to paper as a whole can be seen in Table 1.

Table 1. Relation *Paper* to *Paper*

X / Y	P1	P2	P3	P4	P5	P6
P1	1,00	0,78	0	0	0,64	0,54
P2	0,93	1,00	0	0	0,64	0,73
P3	0	0	1,00	0,97	0,36	0,22
P4	0	0	0,83	1,00	0,29	0,22
P5	0,60	0,50	0,28	0,26	1,00	0,70
P6	0,67	0,75	0,22	0,26	0,93	1,00

3. From the existing data of keywords related to the paper, then we will get the paper on the relationship between keywords. Relationship of paper to keyword can be calculated using formula 2.2 [5]:

$$H_{Dj}(P_j) = \frac{\mu_{pi}(D_j)}{\mu_{pi}(D_1) + \mu_{pi}(D_2) + \ldots + \mu_{pi}(D_m)} \qquad (2.2)$$

   Where:

   R: Relation

   P_i: *Paper*/document i

   P_j: *Paper*/document j

   D: *Keyword*

   μ: *Membership function* as a *mapping* $\mu P_i: D \rightarrow [0,1]$.

   Example: Calculate the weight of *keyword* (d2) for *paper* 1

$$\eta_{Dj}(P_j) = \frac{0.3}{0.3+0.7+1+1} = \frac{0.3}{3} = 0.1$$

   so the final result is:

   $D_1 = (0.25/P_3, 0.32/P_4)$,

   $D_2 = (0.1/P_1, 0.28/P_2, 0.06/P_5, 0.24/P_6)$,

   $D_3 = (0.25/P_3, 0.16/P_4)$,

   $D_4 = (0.28/P_3, 0.26/P_4, 0.36/P_5, 0.27/P_6)$,

   $D_5 = (0.23/P_1, 0.22/P_2, 0.25/P_5, 0.27/P_6)$,

   $D_6 = (0.22/P_3, 0.26/P_4)$,

   $D_7 = (0.33/P_1, 0.22/P_2)$,

   $D_8 = (0.33/P_1, 0.28/P_2, 0.36/P_5, 0.27/P_6)$

4. *Similarity* between 2 *keywords* expressed as a function of R where $R: D \times D \rightarrow [0,1]$ as written in the formula 2.3 [5]:

$$R(D_i,D_j) = \frac{\sum_P \min(\eta_{D_i}(p), \eta_{D_j}(p))}{\sum_P (\eta_{D_i}(p))} \qquad (2.3)$$

   Where:

   R: Relation

   P_i: *Paper*/document i

   P_j: *Paper*/document j

   D: *Keyword*

μ: *Membership function* as a *mapping* $\mu P_i: P \rightarrow [0,1]$.

can be found the relationship between keywords with each other eg.:

$$\text{relation } R(D_1,D_3) = \frac{0.25+0.16}{0.25+0.16} = 1$$

$$\text{relation } R(D_3,D_1) = \frac{0.25+0.16}{0.25+0.32} = 0{,}72$$

Calculation of *keyword to keyword* as a whole can be seen in Table 2.

Table 2. Relationship between *Keyword* to *Keyword*

X / Y	D1	D 2	D 3	D 4	D 5	D 6	D 7	D 8
D 1	1,00	0	1,00	0,44	0	1,00	0	0
D 2	0	1,00	0	0,26	0,64	0	0,58	0,53
D 3	0,72	0	1,00	0,35	0	0,79	0	0
D 4	0,89	0,44	1,00	1,00	0,54	1,00	0	0,51
D 5	0	0,91	0	0,44	1,00	0	0,82	0,86
D 6	0,84	0	0,93	0,41	0	1,00	0	0
D 7	0	0,47	0	0	0,46	0	1,00	0,49
D 8	0	1,00	0	0,54	1,00	0	1,00	1,00

.

3. SEARCH ENGINE APPLICATION

Chapter 3 describes the design of systems that are the basis of the developing the application of scientific journal search engines in this research. Basically there are two main processes in this application; they are the indexing process and searching process.
Indexing process takes the longest runtime when compared with the process of searching. It is caused by a cutting process of a document into a word. Indexing process execution depicted in Figure 1.

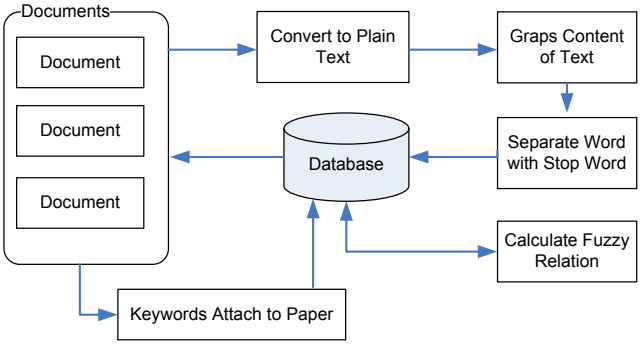

Figure 1. Indexing process

Here is a sequence of description of the indexing process:

1. Retrieve a document that has no plain text file / not yet indexed.
2. Convert the pdf documents into plain text files that are readable by the programming language PHP 5.
3. Input a plain text file data into the database to be stored and through the next process.
4. Read the contents of plain text files and store a long string into an array.
5. Enter a keyword from a document originating from the author to serve as the keyword / main keywords on applications and enter into a database. Keyword is better known by the name of keywords attached to paper.
6. Grab a few sentences from the abstract and titles for snipset when displaying search results on the process of searching.
7. Take the author's name and affiliation to be used in the process of searching with the author and affiliation search mode.
8. Rupture long string that is stored in an array into a word-per-word and separate from the existing stop word. Next save all the words that have been cut into the database and delete its stop word.
9. The last process of this indexing process is the process of calculating the value of fuzzy relations.

- Searching Process

The search engine used 2 kinds of fuzzy methods to process the searching process, both methods are:

a. Ordinary Fuzzy Method
   Ordinary fuzzy method is used when a user input keyword has never been through the process of indexing, but the keywords are included on one or more papers that are stored on the system.
   In this method, the fuzzy value calculation is only carried out on the basis of number of word occurrence; it means the fuzzy value calculation is performed only for papers that have keyword input on its content. This method involves only one fuzzy process has been done before in the indexing process, it is the calculation of keywords to paper fuzzy.
b. Extended Fuzzy Method
   Unlike the ordinary fuzzy method, the calculation by this method not only involves the number of word occurrence alone but also involves all fuzzy calculations have been

done on indexing process. With this method, papers that do not have keyword input on its content will be found.

The use of two methods is for the speed/runtime purpose when the searching process is running. In this search engine, users can not choose the method that will be used, because the system will automatically check the incoming keyword to further select one of two methods for implementing the process of searching. General description of the process of searching can be seen in Figure 2.

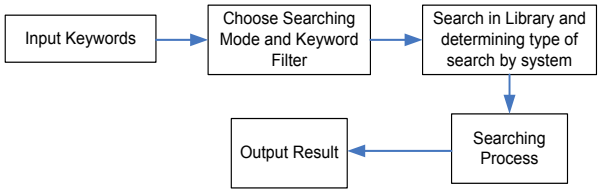

Figure 2. Searching process

4. EXPERIMENTS

In this section, we present an experimental result of new search engine application. This system was built in PHP [1, 2, 3] on a PC with 2.4 GHz Pentium ® 4 CPU and 1 GB of RAM under MS Windows XP Pro.

4.1 Searching Type - *Extended Fuzzy*

Search by extended fuzzy type is a searching type which looks for related papers of related keywords that are input by the user. The implementation of an extended type of fuzzy search with keyword input - one word can be seen in Figure 3.

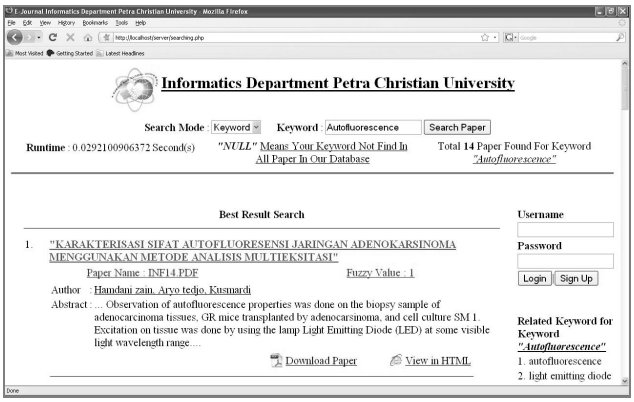

Figure 3. Extended type of fuzzy search with keyword input - one word

The implementation of an extended type of fuzzy search with a keyword input more than one word/phrase can be seen in Figure 4.

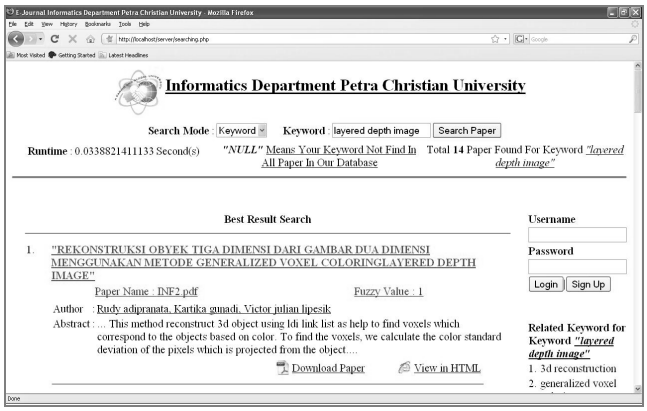

Figure 4. Extended type of fuzzy search with keyword input – more than one word/phrase

4.2 Searching Type - *Ordinary Fuzzy*

Searching by ordinary fuzzy type is a search that only looks for keywords in the paper, inputted by the user. Ordinary type of fuzzy search implementation with a single word keyword input can be seen in Figure 5.

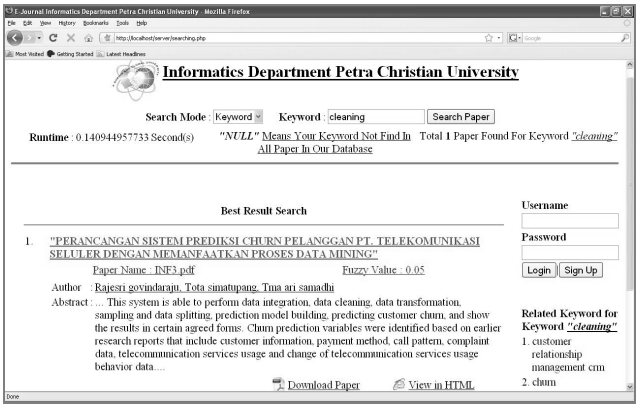

Figure 5. Ordinary type of fuzzy search with a single word keyword input

Ordinary type of fuzzy search implementation with the input of more than one word keywords/phrases can be seen in Figure 6.

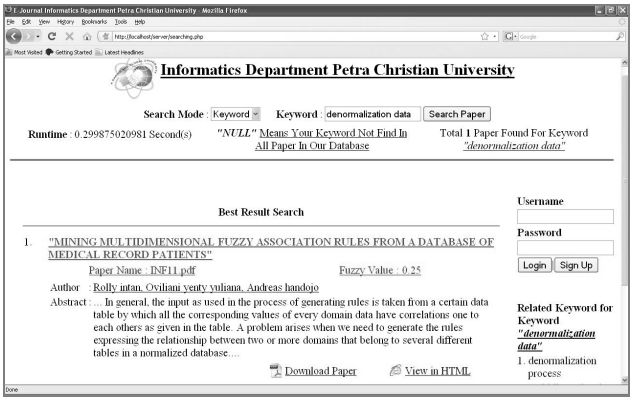

Figure 6. Ordinary type of fuzzy search with more than one work/phrase keyword input

4.3 Search involving symbol and *Stop Word*

In addition to testing with the extended fuzzy search types and ordinary fuzzy, involving a stop word and symbol are also important. This is related to all rules that apply to the program. All these rules are based on assumptions that are not researched before, but just based on observations from the passage of searching module developing. In addition, this test is also a test of all rules that apply to the application.

The Implementation of the search involving symbol and stop word with one-word keywords can be seen in Figure 7.

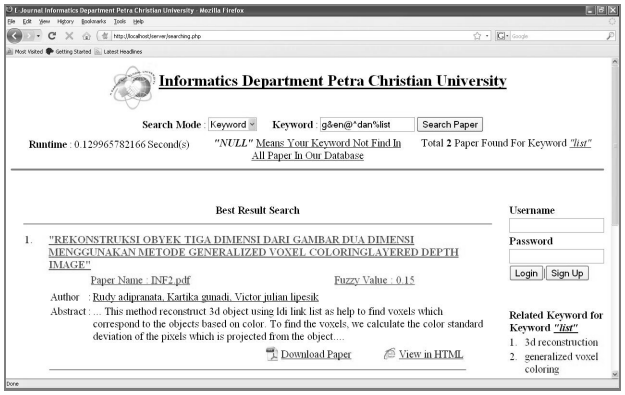

Figure 7. Search involve symbol and stop word with one word keyword input

Search engine implementation which involves symbols and stop word to the keyword input of more than one word/phrases can be seen in Figure 8.

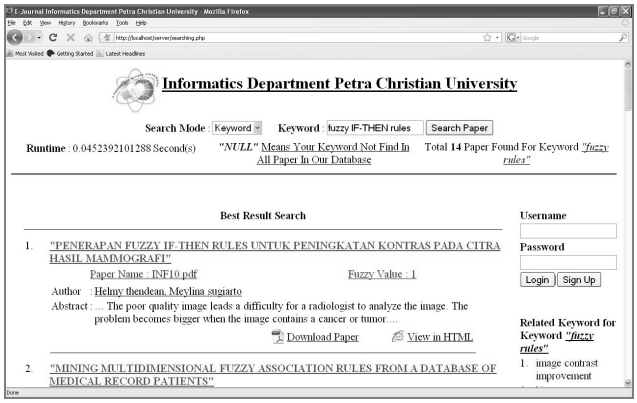

Figure 8. Search engine which involves symbol and *stop word* with more than one word/phrase keyword input

4.4 Searching involves *Keyword Attach to Paper*

Searching involving keyword attached to paper is a searching process performed on the input found on the library tables. This Search use extended fuzzy type because the keyword input is found in the library table. Search engine implementation involves keyword attached to paper with one word keyword input can be seen in Figure 9.

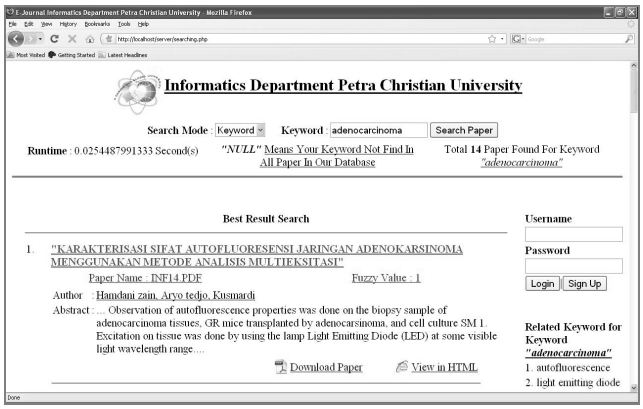

Figure 9. Search engine which involves *keyword attach to paper* with one keyword input

Search engine implementation which involves keyword attach to paper with more than one word/phrases input can be seen in Figure 10.

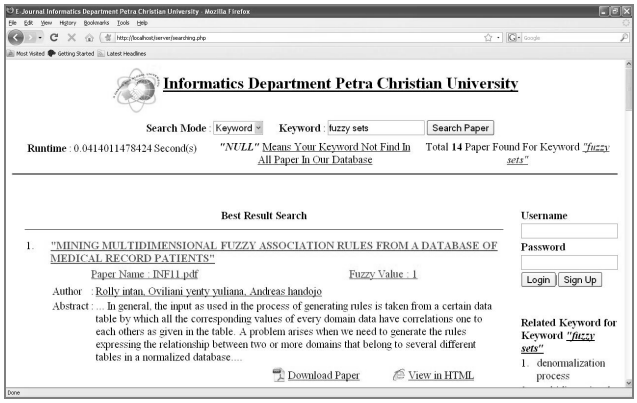

Figure 10. Search engine which involves *keyword attach to paper* with more than one word/phrase keyword input

4.5 Runtime Process

From the experiments conducted, it can be calculated an average speed of the process of searching on search engine applications. The average results of the calculation process of some kind of keyword search can be seen in Table 3.

Table 3. Runtime test result

User inputed Keyword	***Runtime* (second)**
Autofluorescence	0.0254549980
layered depth image	0.0606830120087
Data	0.0613670349121
denormalization *data*	0.343075037003
g&en@^e$rate	0.105587005615
fuzzy IF-THEN rules	0.261485099792

adenocarcinoma	0.0284929275513
fuzzy sets	0.0446968078613
The Average of Runtime	0.8831452369523/8 = **0.1103931546190375**

From 8 keywords input in the system, the average of searching process is 0.1103931546190375 seconds.

Runtime test of indexing process is the main focus of all the testing and experiments performed on the system. From several experiments, it is resulted that more and more number of papers and the number of keywords, the runtime required to complete the indexing process is also getting bigger. From the seventh step in the indexing process, step-core indexing and fuzzy relationship value calculation takes the greatest runtime. The example of execution indexing process is presented in Figure 11.

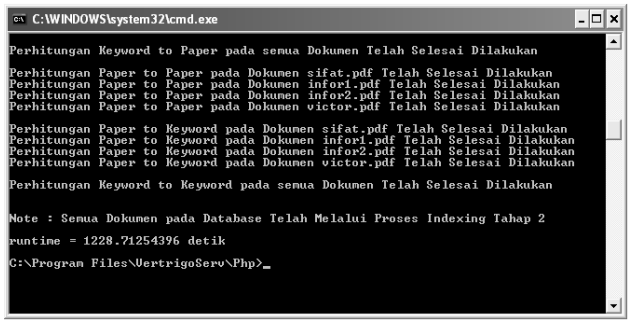

Figure 11. Indexing process

From the testing process it can be described that it takes 1228.71254396 seconds or approximately 20 minutes 5 seconds to index four papers.

5. CONCLUSION

This paper deals with the implementation of fuzzy relation method to support e-journal search engine. Fuzzy value ranking system should be implemented into the calculation of fuzzy relations with respect to the accuracy of produced output.

The Implementation of search engine applications on operating systems other than Windows XP can be done by changing some parts of the segment of the program and replace the use of customized components with the operating system used.

The emphasis of this paper was on feasibility – identification of possible approaches and development of methods to put them into practices.

We are currently working on the implementation process of indexing and searching in a lot of server/multi server. Next, concerns is the quality of the result.

6. REFERENCES

[1] Bibeault, B. and Yehuda, K. 2008. *JQuery in Action*. Manning Publications. USA.

[2] Boronczyk, T., Naramore, E., Gerner, J., Le Scouarnec, Y., and Stolz, J. 2009. *Beginning PHP 6, Apache, and MySQL 6 Web Development*. John Wiley and Sons.

[3] Castagnetto, J., Rawat, H., Schumann, S., Scollo, C., and Veliath, D. 2000. Professional PHP Programming. 3rd Ed. Wrox Press Ltd. USA.

[4] Darmadi, B.A. 2005. *Perancangan dan pembuatan search engine paper/karya ilmiah berbasis web dengan metode fuzzy relation.* Unpublished undergrade thesis, Universitas Kristen Petra, Surabaya.

[5] Intan, R and Masao, M. Toward a fuzzy Thesaurus Based on Similarity in Fuzzy Covering. Australian Journal of Intelligent Information Processing, vol. 8 no. 3, 2004. pp. 132-139.

INFORMATICS DEPARTMENT
PETRA CHRISTIAN UNIVERSITY

CERTIFICATE

This certificate is presented to

LEO WILLYANTO SANTOSO

In recognition of participation in the

2nd International Conference on Soft Computing, Intelligent System and Information Technology (ICSIIT) 2010

Hardrock Hotel, Bali, Indonesia, 1-2 July 2010

Iwan Njoto Sandjaja, MSCS

Program Committee Coordinator